arXiv:1506.02712v2 [quant-ph] 17 Dec 2015

Generation of fresh and pure random numbers for loophole-free Bell tests

Carlos Abellan,[1] Waldimar Amaya,[1] Daniel Mitrani,[1] Valerio Pruneri,[1,2] and Morgan W. Mitchell[1,2]

[1] *ICFO – Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*

[2] *ICREA – Institució Catalana de Recerca i Estudis Avançats, 08015 Barcelona, Spain*

(Dated: December 18, 2015)

We demonstrate extraction of randomness from spontaneous-emission events less than 36 ns in the past, giving output bits with excess predictability below 10^{-5} and strong metrological randomness assurances. This randomness generation strategy satisfies the stringent requirements for unpredictable basis choices in current "loophole-free Bell tests" of local realism [Hensen et al., Nature (London) 526, 682 (2015); Giustina et al., Phys. Rev. Lett. 115, 250401 (2015); Shalm et al., Phys. Rev. Lett. 115, 250402 (2015)].

PACS numbers: 03.65.Ud, 03.65.Ta, 42.50.Ct,

Quantum nonlocality [1] is one of the most striking predictions to emerge from quantum theory. Beyond their fundamental interest, loophole-free Bell tests enable powerful "device independent" information protocols, guaranteed by the impossibility of faster-than-light communication [2]. Bell tests and device-independent protocols employ spacelike separation of measurements to guarantee the nonlocality of correlations [3–8] and the monogamy of correlations under the no-signaling principle [9–11]. To be secure, they must close two space-time loopholes: no basis choice may influence a distant particle (locality loophole), and the entanglement generation must not influence the basis choices (freedom-of-choice loophole). Current efforts [6, 7, 12–14] to simultaneously close the detection [4, 6, 7], locality [3], and freedom-of-choice (FoC) [5, 8] loopholes require random number generators (RNGs) with an unprecedented combination of speed, unpredictability, and confidence [15–17].

Here we combine ultrafast RNG by accelerated laser phase diffusion [18–20] with real-time randomness extraction and metrological randomness assurances [21] to produce a RNGs suitable for loophole-free Bell tests. Because the laser phase diffusion is driven by effects, including spontaneous emission, that are unpredictable both in quantum theory and in an important class of stochastic hidden variable theories, the source can be used to address the "freedom-of-choice" loophole [22]. Using a detailed and validated model of the signal generation process, we show the effectiveness of parity-bit randomness extraction of this source. Under paranoid assumptions, we infer excess predictability below 10^{-5} at 6σ statistical confidence for output based on phase-diffusion events less than 36 ns old. A statistical analysis based on 2.3 Tbits of random data supports the metrological assessment of extreme unpredictability. The results enable definitive nonlocality experiments and secure communications without the need for trusted devices [9, 11, 23, 24].

As shown in Fig. 1, the locality and freedom-of-choice loopholes can be closed by spacelike separation of the random events that determine the basis choice from the distant detection and from the production of the pairs of particles, respectively [10]. This requires generation of randomness in a time window shorter than the light time between the detectors. Closing the "detection loophole" requires high efficiency and motivates protocols very sensitive to predictability of the basis choices. Both experiments employing 100% efficient "event-ready" detection [25] and those employing high-efficiency photodetection [26, 27], are expected to require excess predictabilities ϵ below a few times 10^{-5} [22].

Time and/or frequency metrology, e.g. jitter measurements against stabilized oscillators, are routinely used to determine timing with sub-ns precision and accuracy, allowing reliable identification of spacelike separated events. Achieving similar assurances for unpredictability poses a distinct challenge. For fundamental

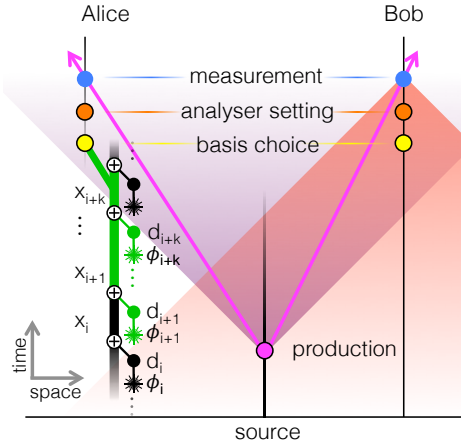

FIG. 1. Space-time diagram for the production of random numbers in a loophole-free Bell scenario. As shown, up to k raw bits can be generated in a time window that is spacelike separated from both (i) the pair generation and (ii) the distant measurement. Laser pulses (✷—) with random phases ϕ_i are converted into raw random bits d_i and extracted bits x_i by a running XOR ($\oplus$) calculation.

reasons, no test on the output of a RNG can demonstrate randomness, and statistical characterization of the RNG process becomes the critical task. Here we develop statistical metrology for a short-delay RNG, in analogy to earlier work with high-throughput RNGs [19, 21, 28]. The excess predictability ϵ is exponentially reduced by randomness extraction (RE) [29]: in real time, we compute the parity of several raw bits to produce one very unpredictable extracted bit for the setting choice.

The RNG and its behavior are illustrated in Fig. 2. A single-mode laser diode (LD) is strongly current modulated, going above threshold for about 2 ns of every 5 ns cycle, to produce a train of optical pulses with very similar waveforms, as seen in Fig. 2 (b). In the time below threshold strong phase diffusion randomizes the optical phase within the laser [30, 31], and, thus, the relative phase ϕ from one pulse to the next. At the time a pulse leaves the laser, it is already a macroscopic ($\sim$ mW) signal, with a phase that has been fully randomized by the microscopic process of spontaneous emission. An unbalanced Mach-Zehnder interferometer (UMZI) converts the train of phase-random pulses into amplitude-random pulses; see Fig. 2 (c). These are detected with a fast photodiode, giving a voltage signal $v(t)$. A fast comparator and a D-type flip-flop digitize (with one-bit resolution) the signal at times t_i to give, at 200 Mbps, raw digital values $d_i = \theta(v(t_i) - v_{\rm ref}(t_i))$, where θ is the Heaviside step function and $v_{\rm ref}$ is the comparator reference level. To correct for drifts in laser power, the reference level is set by feedback from the raw digital values via an integrator with a 1 ms time constant. We observe a raw-bit average of $\langle d\rangle = \frac{1}{2}[1 + 6.9(1)\times 10^{-4}]$.

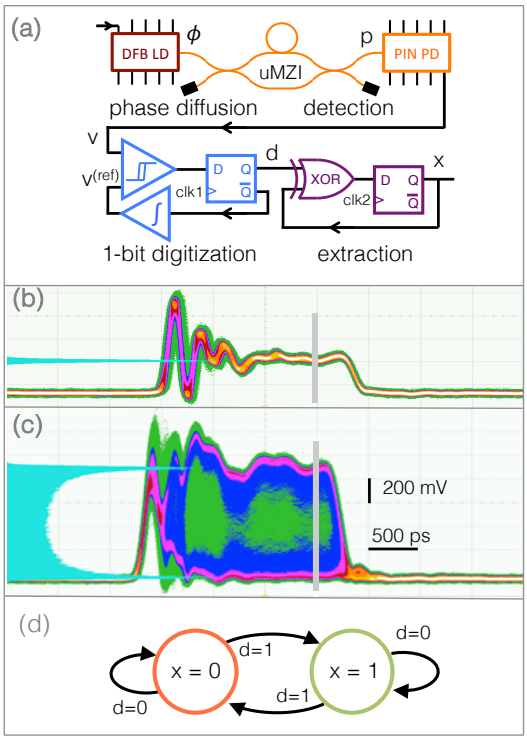

FIG. 2. Random number generation for loophole-free Bell tests. (a) Experimental schematic. Laser pulses with random phases ϕ_i from a distributed feedback laser diode (DFB LD) are converted to random powers p_i by an UMZI and detected with a linear photoreceiver (PIN PD) to give analog voltages v_i. These are one-bit digitized with a comparator and D-type flip-flop to give raw bits d_i, and summed modulo 2 with an XOR gate to give extracted bits x_i. The output value x_{i+k} includes the parity of k raw bits, d_{i+1} to d_{i+k}, due to pulses spacelike separated from the distant measurement and from the entanglement production. (b), (c) "Persistence mode" visualization of $v(t)$ statistics. Warmer colors show greater frequency; teal histogram on the left axis describes the voltages sampled inside the time window indicated in gray. (b) Noninterfering pulses obtained by blocking the long interferometer path; (c) interfering pulses. (d) Two-state machine describing the randomness extraction.

An XOR gate and a second flip-flop perform a running parity calculation, updating the output x as $x_i = x_{i-1} \oplus d_i$, where $\oplus$ indicates addition modulo 2. This describes a two-state machine, see Fig. 2 (d), that changes state every time a new raw bit $d_i = 1$. Note that x accumulates the parity of all preceding raw bits, only k of which will be spacelike separated from the distant measurement. When a bit $x_{i+k} = x_i \oplus d_{i+1} \oplus \ldots \oplus d_{i+k}$ is used for a basis setting, x_i contributes no spacelike separated randomness, and the predictability of x_{i+k} will be determined by $d_{i+1} \oplus \ldots \oplus d_{i+k} \equiv D_{i,k}$. Writing the predictability of d_i, i.e. the probability of the more likely value, as $\mathcal{P}(d_i) = \frac{1}{2}(1+\epsilon_i)$, where $\epsilon_i \geq 0$ is the instantaneous excess predictability, we find (see the Supplemental Material) that if $\epsilon_i \leq \epsilon_{\rm max}$, the predictability of the parity of k bits is bounded as $\mathcal{P}(D_{i,k}) \leq \frac{1}{2}(1+\epsilon_{\rm max}^k)$. The RE output approaches ideal randomness exponentially in k.

We define the "freshness time" to be the interval between the earliest spontaneous-emission events required for randomizing a bit and the bit's availability for use. The largest phase diffusion occurs at the rising edge of the current pulse, when the intracavity photon number is at a minimum [30]. The freshness time for a single bit $\tau_{\rm f}$, measured from a rising edge of the electrical modulation signal to availability of the corresponding bit at the output port is bounded by $10.01 < \tau_{\rm f} < 11.07$ ns with a p-value $< 1.4\times 10^{-6}$ (see the Supplementary Material). Since we can use $(k-1)$ extra bits that are still spacelike separated, $(k-1)$ additional clock cycles of 5 ns each are needed. In total, the freshness time to produce and propagate k bits from the oldest spacelike separated spontaneous-emission event to the output port is $\tau_{\rm f}^{(k)} = \tau_{\rm f} + 5\times(k-1)$.

Metrological assurances proceed from the interference behavior. The instantaneous power reaching the detector is

$$p_I(t) = p_S(t) + p_L(t) + 2\sqrt{p_S(t)p_L(t)}\cos\phi(t), \qquad (1)$$

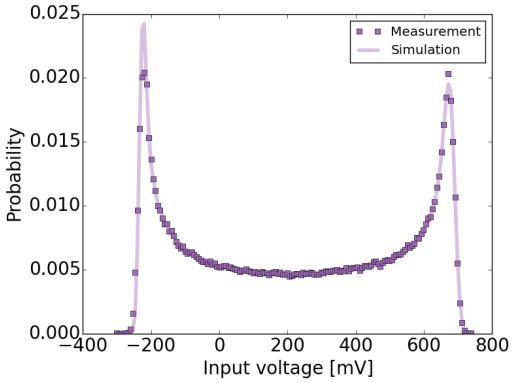

FIG. 3. Histogram (points) of analog signals v, showing an arcsine distribution, and prediction (line) from a Monte Carlo simulation of Eq. (2) using measured rms deviations for all noise sources $v_{\rm S}, v_{\rm L}, v_{\rm PD}, v_{\rm H}$ and $v_{\rm ref}$, and a fitted visibility $\mathcal{V} = 0.955$. Voltage scale is offset due to ac-coupling.

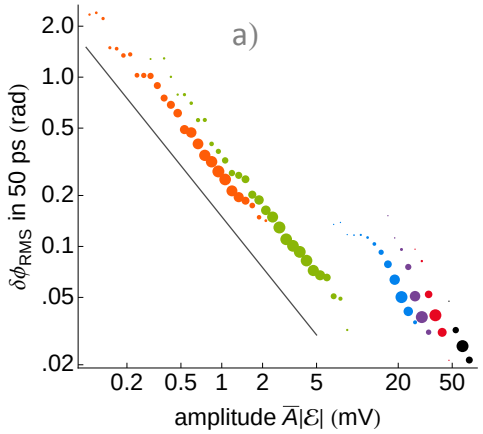

FIG. 4. Observed frequencies (spot size) of field amplitude $|\mathcal{E}|$ and resulting phase dispersion $\delta\phi_{\rm rms}$ over 50 ps, measured by heterodyne detection with gain $\overline{A}$ for continuous currents (colors, left to right) 15, 16, 16.5, 17, 17.5, 19 mA. Grey line shows $\delta\phi_{\rm rms} \propto |\mathcal{E}|^{-1}$ scaling of spontaneous-emission-driven phase diffusion.

where p_S and p_L are the contributions of the short and long paths, respectively. We note that optical visibility is guaranteed by the single-spatial-mode fiber interferometer and the single-longitudinal-mode laser emission. Including detection noise and finite bandwidth effects, the electronic output is (see the Supplemental Material)

$$v(t) = v_{\rm S}(t) + v_{\rm L}(t) + v_{\rm H}(t) + v_{\rm PD}(t) + v_\phi(t), \quad (2)$$

where $v_{\rm S}$ and $v_{\rm L}$ are the short- and long-path contributions, respectively, $v_{\rm H}$ describes "hangover errors," i.e. delayed contributions from earlier pulses [21], and $v_{\rm PD}$ is the detector noise. $v_\phi = 2\mathcal{V}\sqrt{v_{\rm S} v_{\rm L}}\cos\phi$ is the trusted signal from interference, where $\mathcal{V}$ is the visibility after detection. In Fig. 3), we show the distribution of v, sampled at the moment indicated in Fig. 2 (c), and infer $\mathcal{V} \sim 95\%$ using a Monte Carlo simulation of Eq. (2) (see the Supplementary Material). As shown in Fig. 2 (c), we take a sample < 2 ns after the rising edge occurs, chosen late in the pulse so that relaxation oscillations have decayed. The histogram is well modeled by the arcsine distribution, which describes the cosine of a uniformly distributed phase.

The trusted randomness of the signal v originates in ϕ, which between pulses strongly diffuses due to spontaneous emission, as shown in Fig. 4 (see, also, the Supplemental Material). The observable ϕ mod 2π is for all practical purposes uniformly distributed on $[0, 2\pi)$, is unpredictable based on prior conditions, and is independent from one pulse to the next [32], irrespective of any other phase shifts [21].

With the exception of $\cos\phi$, all contributions to $v(t)$ in Eq. (2), and also $v_{\rm ref}$, contain technical noise due to prior conditions that are not spacelike separated from the distant detection. We define the sum of these untrusted contributions $v_{\rm c} \equiv v_{\rm S} + v_{\rm L} + v_{\rm PD} + v_{\rm H} - v_{\rm ref}$ so that $d_i = \theta[v(t_i) - v_{\rm ref}(t_i)] = \theta[v_\phi(t_i) + v_{\rm c}(t_i)]$, with distribution

$$P(d=1) \equiv P_1 = \frac{2}{\pi}\arcsin\sqrt{\frac{1}{2} + \frac{v_{\rm c}}{2\Delta v_\phi}} \quad (3)$$

where $2\Delta v_\phi = 4\mathcal{V}\sqrt{v_{\rm S} v_{\rm L}}$ is the peak-to-peak range of v_ϕ (see the Supplemental Material). The predictability is $\mathcal{P}(d) \equiv \max[P_1, 1-P_1] \equiv \frac{1}{2}(1+\epsilon)$. Bounding the effect of $v_{\rm c}$ on $\mathcal{P}(d)$ will determine $\epsilon_{\rm max}$, the upper bound on ϵ.

The contributors to $v_{\rm c}$ are electronic signals and are directly measured with a 4 GHz oscilloscope (Agilent Infinitum MSO9404A). For example, the variation of $v_{\rm S}$, the signal in the short path of the interferometer, is measured by blocking the signal from the long path. The measurement gives access to the signal $v_{\rm S} + v_{\rm PD} + v_{\rm O}$, as shown in Fig. 2 (b). Note that the measurement of the signal of the short path is not isolated, but superimposed to the noises in the photodetector $v_{\rm PD}$ and the scope $v_{\rm O}$. To obtain the noise from $v_{\rm S}$ only, we have to subtract the contribution from $v_{\rm PD}$ and $v_{\rm O}$, both directly measurable. Statistics of the measurable noise contributions, always sampled at the same point in the pulse, are given in Table S1 of the Supplemental Material.

To combine the noise sources, we consider three levels of distrust of the equipment: "ordinary," "digitizer paranoid" and "fully paranoid." In all cases, the noises are individually described by the measured statistics of Table S1 in the Supplemental Material, but their assumed correlations vary. In ordinary distrust, we make the physically reasonable assumption that the noise sources are uncorrelated. In digitizer paranoid distrust we assume the comparator, the only nonlinear element of the signal chain, chooses $v_{\rm ref}$ in function of the other noises so

as to maximize the predictability. In fully paranoid distrust, we assume that all noise sources are collaborating to maximize predictability. These assumptions lead to normally-distributed v_c with rms deviations σ shown in Table I. Fluctuations in v_c are, in principle, unbounded but rarely exceed a few standard deviations, a situation that is captured by assigning confidence bounds, in this case to $\mathcal{P}(d)$ and $\mathcal{P}(x)$. For example, considering $v_c = |\langle v_c \rangle| + 6\sigma$ as an upper limit, we compute $\epsilon_{\max}$ using Eq. (3). Noise fluctuations will produce a fraction $P_{6\sigma}$ of the raw bits with $\epsilon > \epsilon_{\max}$, where $P_{6\sigma} \approx 2 \times 10^{-9}$. The excess predictability of the extracted bit exceeds $\epsilon^k_{\max}$ at most this often, even assuming maximally correlated raw-bit excess predictability. See the Supplemental Material for details. Values of σ_{v_c}, $\epsilon_{\max}$ and τ_f for different k and distrust levels are shown in Table I, and in Fig. 5.

TABLE I. Noise and predictability for different trust scenarios. All predictabilities are given for a 6σ confidence level. Times in parentheses indicate freshness time $\tau_f^{(k)}$. See text for details.

Distrust level	Noise σ_{v_c}	Excess predictability $\epsilon^4_{\max}$ (26 ns)	Excess predictability $\epsilon^6_{\max}$ (36 ns)
Ordinary	8.6 mV	2.5×10^{-5}	1.3×10^{-7}
Dig. par.	11.7 mV	8.6×10^{-5}	8.0×10^{-7}
Fully par.	14.5 mV	2.0×10^{-4}	2.9×10^{-6}

Although no test of the output can assure randomness, tests can, nonetheless, detect failure to be random. Because of the low computational capacity of physical RNGs, imperfections are expected mostly in low-order correlations. The autocorrelation of the extracted output $\Gamma_x(k) \equiv \langle x_i x_{i+k} \rangle - \langle x_i \rangle^2$ is bounded by $4|\Gamma_x(k)| \leq \epsilon^k_{\max}$ and, thus, drops off in the same way as the excess predictability (see the Supplemental Material). As shown in Fig. 5, the measured $|\Gamma_x(k)|$ approaches zero exponentially in k, and already with $k = 4$ reaches $|\Gamma_x(k)| < 10^{-6}$, the statistical limit with 1 Tbit.

As detailed in the Supplemental Material, we have applied statistical tests DIEHARDER [33], NIST SP800-22 [34], and TESTU01 ALPHABIT battery [35] to strings of extracted data up to 1 Tbit in length. To study a given k, we first generate a distilled string $z_i \equiv x_{ik}$, i.e. $\{z\}$ is a k-fold subsampling of $\{x\}$. We observe that ALPHABIT, designed to test physical RNGs, is as sensitive as other tests and runs much faster. Already with $k = 3$ extraction ALPHABIT finds no significant patterns in ~ 2.3 Tbits of data, organised as one file of 1 Tbit, two files of 500 Gbits, one file of 80 Gbits, and two files of 64 Gbits. We also tested 300 sequences of lengths 1 Mbit, 0.2 Gbits, 0.5 Gbits, 1.0 Gbit for $k = 1, 2, 3$, and 4, respectively, and compare the failure rates to what is expected for an ideal random source.

In conclusion, we have demonstrated a spontaneous-emission-driven random number generator suitable for closing the locality and freedom-of-choice loopholes in a test that also closes the detection loophole. By combining high-speed phase-diffusion RNG, real-time randomness extraction, and metrological guarantees, we have produced extracted bits traceable to spontaneous-emission events less than 36 ns old and with excess predictability $\epsilon \leq 10^{-5}$. Generation of high-quality random bits in narrow time windows enables definitive tests of quantum nonlocality and "device-independent" technologies guaranteed by the no-signaling principle.

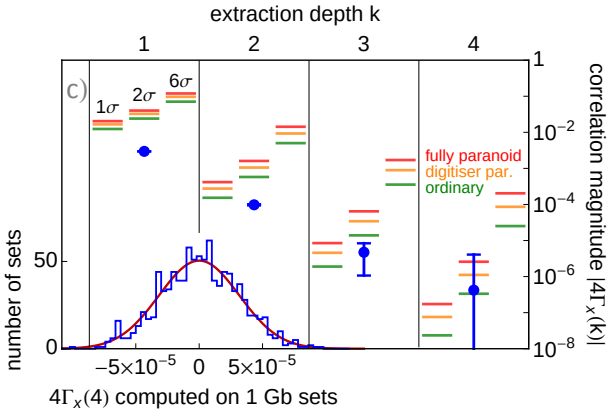

FIG. 5. Measured two-point correlations (blue points) $|4\Gamma_x(k)|$ for extracted bits x, computed on 300 Gbits, showing exponential approach to ideal behavior. Error bars show $\pm 1\sigma$ statistical error in the correlation measurement. Horizontal bars show $\epsilon^k_{\max}$, the metrologically derived upper bound on predictability, for ordinary, (green) digitizer paranoid (orange) and fully paranoid (red) distrust levels and (left to right) 1σ, 2σ, and 6σ confidence. (Lower left) Histogram of $4\Gamma_x(4)$ computed on 1000 runs of 1 Gb each, is normal by the Pearson χ^2 test ($p = 0.27$). Curve shows ideal distribution.

Acknowledgements: We thank M. Giustina, B. Hensen, K. Shalm, J. Kofler, S. Wehner, S. Glancy, S. Jordan, M. Wayne, J. Bienfang, R. Mirin, C. Marquardt, R. Hanson, S.-W. Nam and A. Zeilinger for helpful discussions, and the ICFO electronic workshop for peerless craftsmanship. We thank F. A. C. Diaz-Balart and F. A. C. Smirnov for particularly stimulating conversations. The work was supported by the European Research Council project AQUMET, FET Proactive project QUIC, Spanish MINECO projects MAGO (Ref. FIS2011-23520) and EPEC (FIS2014-62181-EXP), Catalan 2014 SGR 1295, the European Regional Development Fund (FEDER) grant TEC2013-46168-R, and by Fundació Privada CELLEX.

[1] J. S. Bell, Physics **1**, 195 (1964).
[2] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).

[3] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, Phys. Rev. Lett. **81**, 5039 (1998).

[4] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, Nature **409**, 791 (2001).

[5] T. Scheidl, R. Ursin, J. Kofler, S. Ramelow, X.-S. Ma, T. Herbst, L. Ratschbacher, A. Fedrizzi, N. K. Langford, T. Jennewein, and A. Zeilinger, Proc. Natl. Acad. Sci. U.S.A. **107**, 19708 (2010).

[6] M. Giustina, A. Mech, S. Ramelow, B. Wittmann, J. Kofler, J. Beyer, A. Lita, B. Calkins, T. Gerrits, S. W. Nam, R. Ursin, and A. Zeilinger, Nature (London) **497**, 227 (2013).

[7] B. G. Christensen, K. T. McCusker, J. B. Altepeter, B. Calkins, T. Gerrits, A. E. Lita, A. Miller, L. K. Shalm, Y. Zhang, S. W. Nam, N. Brunner, C. C. W. Lim, N. Gisin, and P. G. Kwiat, Phys. Rev. Lett. **111**, 130406 (2013).

[8] C. Erven, E. Meyer-Scott, K. Fisher, J. Lavoie, B. L. Higgins, Z. Yan, C. J. Pugh, B. J.-P., R. Prevedel, L. K. Shalm, L. Richards, N. Gigov, R. Laflamme, G. Weihs, T. Jennewein, and K. J. Resch, Nat. Photonics **8**, 292 (2014).

[9] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Nature (London) **464**, 1021 (2010).

[10] R. Colbeck and R. Renner, Nat. Phys. **8**, 450 (2012).

[11] H.-K. Lo, M. Curty, and B. Qi, Phys. Rev. Lett. **108**, 130503 (2012).

[12] National Institutes of Standards and Technology, "NIST Randomness Beacon," (2011), http://www.nist.gov/itl/csd/ct/nist_beacon.cfm.

[13] J. Hofmann, M. Krug, N. Ortegel, L. Gérard, M. Weber, W. Rosenfeld, and H. Weinfurter, Science **337**, 72 (2012).

[14] H. Bernien, B. Hensen, W. Pfaff, G. Koolstra, M. S. Blok, L. Robledo, T. H. Taminiau, M. Markham, D. J. Twitchen, L. Childress, and R. Hanson, Nature (London) **497**, 86 (2013).

[15] B. Wittmann, S. Ramelow, F. Steinlechner, N. K. Langford, N. Brunner, H. M. Wiseman, R. Ursin, and A. Zeilinger, New J. Phys. **14**, 053030 (2012).

[16] N. Brunner, A. B. Young, C. Hu, and J. G. Rarity, New J. Phys. **15**, 105006 (2013).

[17] J. Kofler, M. Giustina, J.-Å. Larsson, and M. W. Mitchell, ArXiv e-prints (2014), arXiv:1411.4787v3 [quant-ph].

[18] M. Jofre, M. Curty, F. Steinlechner, G. Anzolin, J. P. Torres, M. W. Mitchell, and V. Pruneri, Opt. Express **19**, 20665 (2011).

[19] C. Abellán, W. Amaya, M. Jofre, M. Curty, A. Acín, J. Capmany, V. Pruneri, and M. W. Mitchell, Opt. Express **22**, 1645 (2014).

[20] Z. L. Yuan, M. Lucamarini, J. F. Dynes, B. Fröhlich, A. Plews, and A. J. Shields, Appl. Phys. Lett. **104**, 261112 (2014).

[21] M. W. Mitchell, C. Abellan, and W. Amaya, Phys. Rev. A **91**, 012314 (2015).

[22] J. Kofler, M. Giustina, J.-Å. Larsson, and M. W. Mitchell, ArXiv e-prints (2014), arXiv:1411.4787 [quant-ph].

[23] Y. Liu, T.-Y. Chen, L.-J. Wang, H. Liang, G.-L. Shentu, J. Wang, K. Cui, H.-L. Yin, N.-L. Liu, L. Li, X. Ma, J. S. Pelc, M. M. Fejer, C.-Z. Peng, Q. Zhang, and J.-W. Pan, Phys. Rev. Lett. **111**, 130502 (2013).

[24] A. Mizutani, K. Tamaki, R. Ikuta, T. Yamamoto, and N. Imoto, Sci. Rep. **4** (2014).

[25] B. Hensen, H. Bernien, A. E. Dreau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, Nature (London) **526**, 682 (2015).

[26] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, Phys. Rev. Lett. **115**, 250401 (2015).

[27] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, Phys. Rev. Lett. **115**, 250402 (2015).

[28] F. Xu, B. Qi, X. Ma, H. Xu, H. Zheng, and H.-K. Lo, Opt. Express **20**, 12366 (2012).

[29] S. P. Vadhan, Found. Trends Theor. Comput. Sci. **7**, 1 (2011).

[30] C. H. Henry, IEEE J. Quantum Electron. **18**, 259 (1982).

[31] Y.-Q. Nie, L. Huang, Y. Liu, F. Payne, J. Zhang, and J.-W. Pan, Rev. Sci. Instrum. **86**, 063105 (2015).

[32] This applies not only in a quantum description of laser operation, but in any theory with unpredictable laser phase diffusion. See Supplementary Materials.

[33] R. G. Brown, "Dieharder: A random number test suite," (2004), http://www.phy.duke.edu/∼rgb/General/dieharder.php.

[34] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray, S. Vo, and L. E. Bassham III, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, Tech. Rep. 800-22 (National Institutes of Standards and Technology, 2010) http://csrc.nist.gov/publications/PubsSPs.html#800-22.

[35] P. L'Ecuyer and R. Simard, ACM Trans. Math. Softw. **33**, 22 (2007)

Supplemental material for "Generation of fresh and pure random numbers for loophole-free Bell tests"

DEVICE CONSTRUCTION

The optoelectronic components (LD and PD) are commercial telecommunications devices designed for $\geq$ 10 Gbps direct-modulation data transmission at 1550 nm. The LD is temperature controlled with a Peltier element to ensure wavelength stability and incorporates an optical isolator to prevent optical feedback. The PD incorporates a low-noise trans-impedance amplifier with a linear response. The interferometer is built of polarisation-maintaining single-mode fibre with fibre lengths cut to make the delay of the uMZI 5 ns, equal to the pulse repetition period. The logic elements (comparator, flip-flops and exclusive-OR) are designed for $\geq$ 10 GHz operation with $\leq$ 5 ps jitter. A programmable clock distribution integrated circuit is used to adjust timings with 10 ps resolution to ensure the relative timing of sampling and logic operations. Printed circuit boards were designed, assembled and tested by the ICFO electronic workshop.

SPONTANEOUS EMISSION DRIVEN PHASE DIFFUSION

Laser phase diffusion (LPD) has been intensively studied in LDs, where it is responsible for the free-running line-width [30]. LPD in LDs is driven by spontaneous emission, spontaneous carrier recombination, Johnson noise in the current supply and technical noise sources such as environment-induced current fluctuations. Spontaneous emission contributes a delta-correlated Langevin force [36, 37], which when integrated makes a contribution to ϕ that is independent from one pulse to the next. In the 5 ns period considered here, this contribution is sufficiently large that the distribution of ϕ wraps several times around the phase circle [19, 20].

To validate this model of phase diffusion, we perform heterodyne detection, beating the tested laser against a second "local oscillator" (LO) laser, both running at constant injection current. The beat note, tuned to $\sim$ 3 GHz by temperature adjustment of the LO, is detected with a fast photodiode (ThorLabs DET08CFC, 5 GHz bandwidth), and digitized on a fast oscilloscope (Keysight/Agilent Infiniium MSO9404A) with bandwidth 4 GHz and sampling rate 20 GSa/s. The observed signal is

$$\begin{aligned} v(t) &\propto |\mathcal{E}_{\mathrm{LO}}(t) + \mathcal{E}(t)|^2 \\ &= |\mathcal{E}_{\mathrm{LO}}(t)|^2 + \mathcal{E}_{\mathrm{LO}}(t)\mathcal{E}^*(t) + \mathcal{E}^*_{\mathrm{LO}}(t)\mathcal{E}(t) + |\mathcal{E}(t)|^2 \\ &= \overline{A}^2 + 2\delta A(t) + \overline{A}\left[\cos\Omega t\,\mathrm{Re}\,\mathcal{E}(t) + \sin\Omega t\,\mathrm{Im}\,\mathcal{E}(t)\right] \\ &\quad + O(\delta A^2) + O(\delta A|\mathcal{E}|) + O(|\mathcal{E}|^2), \end{aligned} \tag{S1}$$

where the last three terms are small. Here $A(t) \equiv \overline{A} + \delta A(t) \equiv |\mathcal{E}_{\mathrm{LO}}(t)|$ describes the slowly-varying LO strength, Ω is the angular frequency of the LO in a frame rotating at the mean test laser frequency, and $\mathcal{E}(t) = |\mathcal{E}(t)|\exp[i\phi(t)]$ is the test laser field, also in the rotating frame. Linearising the model by dropping the last three terms, we extract $\delta A(t), \mathrm{Re}\,\mathcal{E}(t)$ and $\mathrm{Im}\,\mathcal{E}(t)$ by Rauch-Tung-Striebel smoothing, i.e., by bi-directional Kalman filtering [38, 39].

As seen in Fig. S1, the signals show direct evidence for phase-diffusion by spontaneous emission: at large or moderate amplitudes, the phase is stable for several cycles. It makes rapid changes, however, either increasing or decreasing the phase, when the amplitude $|\mathcal{E}|$ becomes small. In contrast to a laser above threshold, which diffuses in phase but returns to a finite equilibrium amplitude, in a laser below threshold $\mathcal{E}$ executes a diffusive motion about the point $\mathcal{E} = 0$. Because they contribute to the field with random phases, spontaneous emission events drive diffusion equally in the real and imaginary components of $\mathcal{E}$, giving a phase diffusion coefficient inversely proportional to the amplitude [36, 37]. This behavior is quantified in Fig. 4, which shows statistics acquired from 800 ns-long traces of signals of the kind shown in Fig. S1.

We quantify the phase diffusion using the Holevo variance $\mathrm{var}_{\mathrm{H}}(\phi) \equiv |\langle\exp[i\phi]\rangle|^{-2} - 1$, which is well-behaved for cyclic variables and approaches the ordinary variance $\mathrm{var}(\phi) \equiv \langle\phi^2\rangle - \langle\phi\rangle^2$ for small values. For the data shown in Fig. 4, we choose a time separation $\Delta t = 50$ ps, corresponding to the oscilloscope sampling period, and compute the phase diffusion as

$$\delta\phi_{\mathrm{RMS}} \equiv \sqrt{|\langle e^{i[\phi(t+\Delta t)-\phi(t)]}\rangle_{|\mathcal{E}|}|^{-2} - 1}, \tag{S2}$$

where the angle brackets $\langle\cdot\rangle_{|\mathcal{E}|}$ indicate the average over all times t with a given $|\mathcal{E}|$. The scaling $\delta\phi_{\mathrm{RMS}} \propto |\mathcal{E}|^{-1}$ observed below threshold is a direct confirmation that spontaneous emission is responsible for the observed strong phase diffusion. Phase changes due to refractive index variation would cause $|\mathcal{E}|$-independent phase changes and fluctuating nonlinearities would give diffusion increasing with $|\mathcal{E}|$.

The 18 mA current used in Fig. S1 is near the lasing threshold, resulting in multi-ns periods of slow diffusion. In contrast, when used in the RNG, the laser is taken far below threshold to produce much faster phase diffusion and maintained in this condition for 3 ns to achieve a full spontaneous emission driven phase randomization.

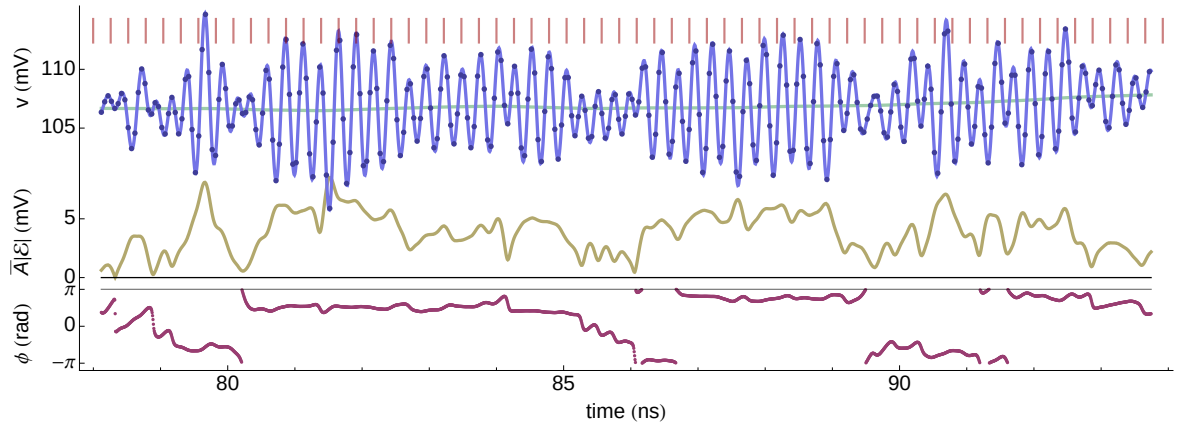

FIG. S1. Direct observation of spontaneous emission driven phase diffusion in the DFB laser at a constant, threshold-region drive current of 18 mA. Heterodyne signal $v(t)$ (top, blue): Points show 20 Gsa/s oscilloscope data, curve shows prediction of Kalman filter, from which amplitude $\overline{A}|\mathcal{E}|$ (middle, beige), phase ϕ (bottom, maroon) and baseline A^2 (top, green) are obtained. Ruler (top, red) is a guide to make phase changes more visible. As expected for phase diffusion by spontaneous emission, rapid phase changes occur when the amplitude dips toward zero.

UNPREDICTABILITY OF PHASE DIFFUSION IN NON-QUANTUM THEORIES

Bell inequalities test local realistic hidden-variable theories (HVTs), and must be analyzed under the assumptions of local realism, not those of quantum mechanics. Spontaneous emission and laser phase diffusion are observable phenomena and unlike, say, entanglement, do not in themselves belong to any particular theory. Moreover, it is an experimental observation, repeated on many kinds of lasers, that the phase of a laser executes a diffusive motion proportional to the spontaneous emission rate. Meanwhile, spontaneous emission, by Einstein's thermodynamic A and B coefficient argument, is a necessary accompaniment of stimulated emission, and thus of laser amplification [40]. It would thus be difficult to exclude spontaneous emission, the archetype of a stochastic physical process, from the description of laser phase diffusion.

To describe the role of randomness in a Bell test, we first note that any given experiment can only test limited classes of HVTs. Fully deterministic HVTs, which Bell named "superdeterminism," cannot be tested by any Bell test, because in superdeterminism it is impossible to make free choices for the settings [22] [41]. In contrast, stochastic HVTs, i.e., those in which some events are unpredictable even in principle, can be tested. Within this class, different HVTs will hold that some subset of: spontaneous emission, chaotic evolutions, thermal fluctuations, human decision-making, and any number of other arguably unpredictable processes, are stochastic. A Bell test can exclude some of these classes of stochastic HVTs.

Directly observable and thus theory-independent contributors to laser phase diffusion include spontaneous emission, Johnson noise in the injection current, and carrier density fluctuation due to spontaneous carrier recombination. Because the phase is cyclic, if any one of these is sufficient to randomize the phase, it remains a pure random variable even when the others are known in advance. A test using these generators can thus exclude HVTs in which any of these processes is stochastic. In addition, a classical HVT in which the laser action is produced by classical electrons moving inside a the laser gain material would almost certainly be chaotic, allowing the test of another class of stochastic HVTs.

DETECTION MODEL

When detected, $p_I(t)$ gives rise to an analog voltage signal $v(t) = [h * p_I](t) + v_{\rm PD}(t)$, where $h(t)$ is the impulse response of the detection system, $*$ indicates convolution, and $v_{\rm PD}$ is detector noise. Because the photoreceiver has a bandwidth much larger than the pulse repetition frequency, the signal mostly represents optical energy received within the last $\tau_{\rm PD} \sim 100\,$ps, and thus from the present pulse. Nevertheless, we need to take account of "hangover error," i.e., delayed signal from previous pulses. We divide the impulse response as $h(t) = h_{\rm f}(t) + h_{\rm s}(t)$, where $h_{\rm f}(t)$ describes the short-time response of the detection system and is nonzero only for $0 \le t \le 5\,$ns, i.e. for one pulse repetition period, whereas $h_{\rm s}(t)$ is nonzero only for $t > 5\,$ns and describes the delayed response. We define $v_{\rm S} \equiv h_{\rm f} * p_S$, $v_{\rm L} \equiv h_{\rm f} * p_L$, $v_{\rm H} \equiv h_{\rm s} * p_I$ and the visibility $\mathcal{V}$ through $v_\phi \equiv \mathcal{V}\cos\phi\sqrt{v_{\rm S}v_{\rm L}} \equiv h_{\rm f} * (\cos\phi\sqrt{p_S p_L})$. We then have

$$\begin{aligned} v &= h_{\rm f} * p_I + h_{\rm s} * p_I + v_{\rm PD} \\ &= v_{\rm S} + v_{\rm L} + 2\mathcal{V}\sqrt{v_{\rm S}v_{\rm L}}\cos\phi + h_{\rm s} * p_I + v_{\rm PD} \\ &= v_{\rm S} + v_{\rm L} + v_\phi + v_{\rm H} + v_{\rm PD}. \end{aligned} \tag{S3}$$

Including the reference level noise $v_{\rm ref}$, the total untrusted noise is $v_{\rm c} \equiv v_{\rm S} + v_{\rm L} + v_{\rm PD} + v_{\rm H} - v_{\rm ref}$.

It is natural to ask whether all relevant noises are included in $v_{\rm c}$. We note that, with the exception of $v_{\rm S}$ and $v_{\rm L}$, the individual noise contributions are each defined to be the total noise arising in some part of the conversion from optical to analog to digital. The noise in converting from optical to analog is $v_{\rm H} + v_{\rm PD}$. v_H is defined as all the fluctuation that comes from prior pulses due to slow response of the detection system, whereas v_D is defined as all other classical noise arising in the detection process. The noise in converting from analog to digital is $v_{\rm ref} - \langle v_{\rm ref}\rangle$. Again, this is defined to be all noise of that kind.

$v_{\rm S} - \langle v_{\rm S}\rangle$ and $v_{\rm L} - \langle v_{\rm L}\rangle$ describe fluctuations in the individual pulse strengths as they leave the laser. The fact that there are only two such sources comes from the fact that we are using an interferometer with two paths (short and long). This describes the topology of the interferometer, and is unambiguous.

Distributions of analog and digital signals

The digitized signal $v = v_\phi + v_{\rm c}$ is the sum of $v_\phi = \Delta v_\phi \cos\phi$, where ϕ is trusted to be fully random and independent, and $v_{\rm c}$, which is untrusted. v_ϕ is described by an arcsine distribution, with probability density function (PDF)

$$P(v_\phi = x) = \frac{1}{\pi\sqrt{\Delta v_\phi - x}\sqrt{x + \Delta v_\phi}} \tag{S4}$$

if $|x| < \Delta v_\phi$ and zero otherwise. The cumulative distribution function is

$$P(v_\phi \le x) \equiv \int_{-\infty}^{x} dx\, P(x) = \frac{2}{\pi}\arcsin\sqrt{\frac{1}{2} + \frac{x}{2\Delta v_\phi}}, \tag{S5}$$

for $|x| \le \Delta v_\phi$. We assume $v_{\rm c}$ is available to the distant particle and ask what distribution the particle would predict for v, knowing $v_{\rm c}$. This describes a conditional distribution $P(v = x|v_{\rm c})$ equal to $P(v_\phi = x - v_{\rm c})$, which is simply $P(v_\phi = x)$ shifted by $v_{\rm c}$. When digitized as $d = \theta[v_\phi + v_{\rm c}]$, where θ is the Heaviside function, the conditional distribution for d, i.e. the probability mass function, is

$$\begin{aligned} P(d=0) &= P(v_\phi + v_{\rm c} \le 0) = P(v_\phi \le -v_{\rm c}) \\ &= \frac{2}{\pi}\arcsin\sqrt{\frac{1}{2} - \frac{v_{\rm c}}{2\Delta v_\phi}}, \end{aligned} \tag{S6}$$

for $|v_{\rm c}| \le \Delta v_\phi$. $P(d=1)$ is then $1 - P(d=0) = \frac{2}{\pi}\arcsin\sqrt{\frac{1}{2} + \frac{v_{\rm c}}{2\Delta v_\phi}}$ as given in Eq. (3).

RANDOMNESS EXTRACTION

Consider two partially-random bits $a = \theta[v_\phi^{(a)} + v_{\rm c}^{(a)}]$ and $b = \theta[v_\phi^{(b)} + v_{\rm c}^{(b)}]$ generated as above, where superscripts $^{(a)}$ and $^{(b)}$ indicate distinct realizations of the variables. Because $v_{\rm c}^{(a)}$, $v_{\rm c}^{(b)}$ are not trusted to be independent, the joint probability $P(a = x, b = y)$ is not in general separable. Nevertheless, the conditional joint probability separates: $P(a = x, b = y|v_{\rm c}^{(a)}, v_{\rm c}^{(b)}) = P(a = x|v_{\rm c}^{(a)})P(b = y|v_{\rm c}^{(b)})$, because the conditioned a and b are a function only of the independent $v_\phi^{(a)}$ and $v_\phi^{(b)}$, respectively. This allows us to use the properties of independent random variables in computing predictability bounds.

For a partially-random bit d, the predictability is $\mathcal{P}(d) \equiv \max[P(d=0), P(d=1)]$. An ideal random bit has predictability 1/2, and writing $\mathcal{P}(d) = \frac{1}{2}(1 + \epsilon_d)$ we find the error $\epsilon_d = 2\mathcal{P}(d) - 1$. Considering two partially-random bits a and b, with predictabilities $\mathcal{P}(a), \mathcal{P}(b)$, respectively, $\mathcal{P}(a\oplus b)$, the predictability of $a\oplus b$ is the larger of

$$P(a \oplus b = 0) = P(a=0)P(b=0) + P(a=1)P(b=1)$$

and

$$P(a \oplus b = 1) = P(a=0)P(b=1) + P(a=1)P(b=0),$$

from which we find

$$\mathcal{P}(a \oplus b) = \mathcal{P}(a)\mathcal{P}(b) + [1 - \mathcal{P}(a)][1 - \mathcal{P}(b)] \tag{S7}$$

or $2\mathcal{P}(a \oplus b) - 1 = [2\mathcal{P}(a) - 1][2\mathcal{P}(b) - 1]$, so that

$$\epsilon_{a\oplus b} = \epsilon_a \epsilon_b. \tag{S8}$$

Moreover, if $\epsilon_a \le \epsilon_{a,\max}$ and $\epsilon_b \le \epsilon_{b,\max}$ i.e. if the errors are bounded from above, then

$$\epsilon_{a\oplus b} \le \epsilon_{a,\max}\epsilon_{b,\max} \tag{S9}$$

because $\epsilon_{a\oplus b}$ is monotonically increasing in both ϵ_a and ϵ_b. When computing $x_{i+k} \equiv x_i \oplus d_{i+1} \oplus \ldots \oplus d_{i+k}$, where $\epsilon_{d_j} \le \epsilon_{d_j,\max}$, repeated application of Eq. (S9) gives

$$\epsilon_{x_{i+k}} \le \Pi_{j=1}^{k}\epsilon_{d_{i+j},\max}. \tag{S10}$$

This shows an exponential approach to ideal randomness. Note that x_i, which is due to events in the past lightcone of the detection, as shown in Fig. 1, contributes no randomness and $\epsilon_{x_i} = 1$.

NOISE MEASUREMENTS

Statistics are measured with an AC-coupled oscilloscope (Agilent Infinitum MSO9404A) with 4 GHz input bandwidth and 8 bit resolution. Statistics are acquired

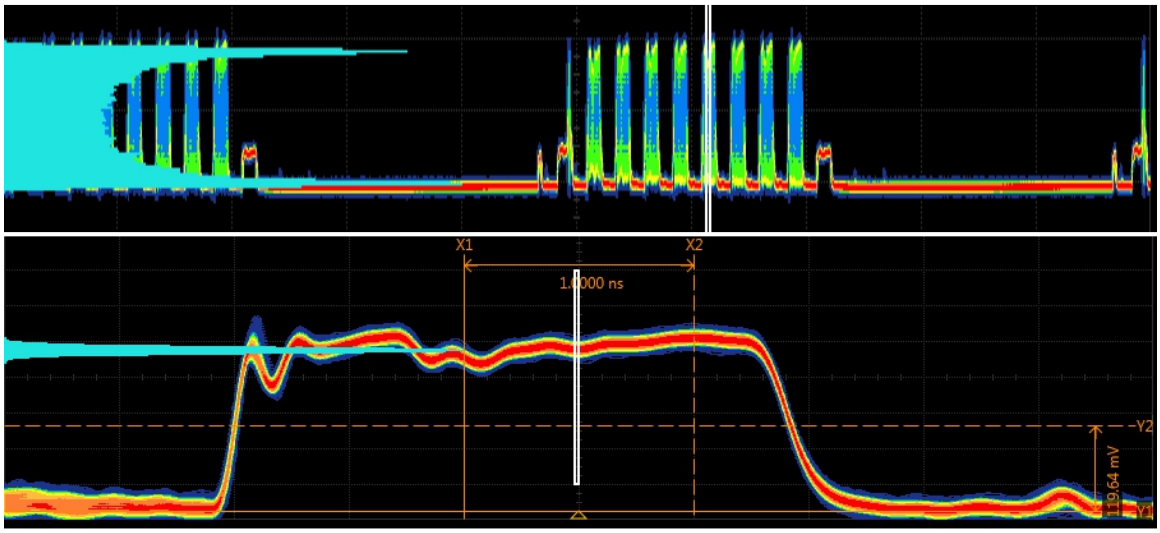

FIG. S2. Measurement of "hangover error." Upper image shows persistence mode oscilloscope trace of trains of 9 optical pulses giving rise to 8 strongly interfering pulses followed by one pulse without interference, shown in higher resolution in the lower image. Pulse repetition period is 5 ns. Teal histograms on left vertical axis show data sampled in the region defined by the white cursors. "Hangover," i.e., remaining variation from previous pulses, is visible at the start of the lower trace, and decreases approaching the sampled point.

as histograms of sampled voltages within a 50 ps window, as shown in Fig. 2 b) & c). All signals in Table S-I except $v_{\rm O}$ and $v_{\rm ref}+v_{\rm O}$ were measured at the photodiode output. To measure $v_{\rm L}+v_{\rm PD}+v_{\rm O}$ or $v_{\rm S}+v_{\rm PD}+v_{\rm O}$ we block the short or long path of the interferometer, respectively. To measure hangover errors, we use a fast analog switch (Mini-circuits ZASWA-2-50DR+) to periodically block the train of current pulses to the LD, see Section . Immediately following the turn-off, a single long-path pulse arrives to the photodiode with no corresponding short-path pulse, thus producing a voltage $v_{\rm L}+v_{\rm H}+v_{\rm PD}+v_{\rm O}$. The full interferometer produces $v_\phi+v_{\rm S}+v_{\rm L}+v_{\rm PD}+v_{\rm H}+v_{\rm O}$.

TABLE S-I. Measured noise statistics. All voltages in mV. $v_{\rm O}$, $v'_{\rm O}$ and $v''_{\rm O}$ indicate the oscilloscope noise at gains of 50 mV/division, 100 mV/division and 200 mV/division, respectively. $v_{\rm L}+v_{\rm H}+v_{\rm PD}+v_{\rm O}$ is measured using an interrupted pulse train, as in Section . $v_{\rm ref}+v_{\rm O}$ is measured using the x-y method of Sec. . Δv_ϕ is determined from the fit of Fig. 3.

measured variables	r.m.s. dev.	derived variables	mean	r.m.s. dev.
$v_{\rm O}$	1.7			
$v_{\rm PD}+v_{\rm O}$	1.9	$v_{\rm PD}$		0.8
$v_{\rm S}+v_{\rm PD}+v_{\rm O}$	2.2	$v_{\rm S}$	251	1.1
$v_{\rm L}+v_{\rm PD}+v_{\rm O}$	2.3	$v_{\rm L}$	251	1.3
$v_{\rm L}+v_{\rm H}+v_{\rm PD}+v_{\rm O}$	4.4	$v_{\rm H}$		3.8
$v'_{\rm O}$	3.4			
$v_{\rm ref}+v'_{\rm O}$	8.4	$v_{\rm ref}$		7.7
$v''_{\rm O}$	7			
$v_\phi+v_{\rm S}+v_{\rm L}+v_{\rm PD}+v_{\rm H}+v''_{\rm O}$	354	Δv_ϕ	483	

The reference voltage $v_{\rm ref}$ is studied in two ways. 1) We measure the width of the comparator transition threshold, by correlating the digital output to the random analog input (see subsection). 2) direct measurement of $v_{\rm ref}$ with the oscilloscope. These two methods give consistent results for $\sigma_{v_{\rm ref}}$. In contrast to v_ϕ, all technical noises contributing to v appear normally distributed with histograms obeying the "68-95-99.7" rule.

We simulated Eq. (2) by generating normally distributed random numbers with means and r.m.s. deviations given by the measurements shown in Table S-I. The phase was simulated to be fully randomised. The resulting distribution was fitted to an arcsine distribution, as shown in Eq. (S4), finding $\langle 2\Delta v_\phi\rangle = 966\,{\rm mV}$. The fit is shown in Fig. 3. In calculating the 6σ bound on $\mathcal{P}$, we adjust this value downward by a multiplicative factor $\sqrt{1-6\sigma_{v_{\rm S}}/\langle v_{\rm S}\rangle}\sqrt{1-6\sigma_{v_{\rm L}}/\langle v_{\rm L}\rangle}\approx 0.971$ to conservatively account for fluctuations in Δv_ϕ.

We find $\langle v_{\rm c}\rangle$, the mean of the combined noises, from $\langle v_{\rm c}\rangle = 2\Delta v_\phi[\langle\sin^2(P_1\pi/2)\rangle - 1/2] \approx 2\Delta v_\phi[\sin^2(\langle P_1\rangle\pi/2)-1/2] = 3.0\,{\rm mV}$, using the measured $\langle P_1\rangle = 0.50035$. The approximation of $\langle\sin^2(P_1\pi/2)\rangle$ is justified in light of Eq. (3) and the observed $|v_{\rm c}| \ll \Delta v_\phi$.

Measurement of hangover errors

To measure the hangover errors, we periodically interrupt the modulation of the LD using an RF switch (Mini-Circuits ZASWA-2-50DR+) at 10 MHz. This generates a train of optical pulses at the output of the laser. Due to the relative path difference in the interferometer, three

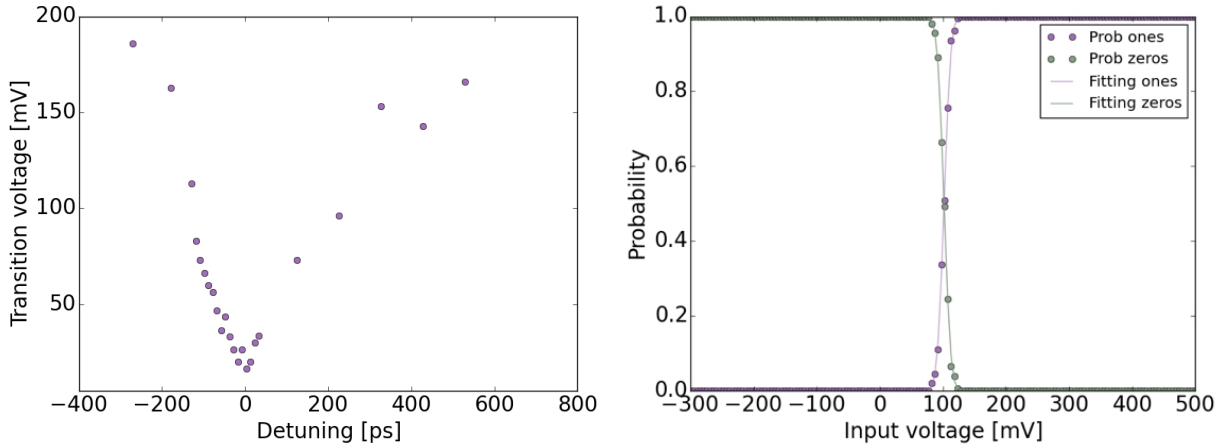

FIG. S3. (Left) Probability of getting an output zero or one for each input analog voltage v_2. As observed, there is a small region near the threshold region in which input values can give zero or one with certain probabilities. The transition probability from $0 \to 1$ ($1 \to 0$) is in excellent agreement with the integral of a Gaussian with $\sigma_{v_{\rm ref}} \leq 7.7\,\mathrm{mV}$. (Right) Measured noise as a function of the detuning between the oscilloscope sampling time and the flip-flop sampling time. We emphasise this measurement is conservative, as the measured error is necessarily larger than the real error.

different types of pulses emerge: (i) the first output pulse, which contains only a short-path contribution and experiences no interference, (ii) the intermediate pulses, which contain both short and long-path contributions and show interference and (iii) the last pulse, which contains only a long-path contribution and thus shows no interference. This last pulse also contains any delayed response, i.e. "hangover," from previous pulses. By measuring the statistical behavior of this last pulse, and comparing against the long-path signal obtained by blocking the short-path, we can recover the contribution from hangover errors. This is illustrated in Fig. S2, which shows a train of nine optical pulses. For the data reported in Table S-I trains of ten pulses were used.

Noise in the reference level: Input-Output analysis

In addition to a direct measurement of $v_{\rm ref}$, the comparator reference voltage, we also measured the input-output relation, which quantifies the performance of the 1-bit analog-to-digital conversion (ADC). Note that while an ideal comparator converts each analog input into an unambiguous digital output, a real comparator has noise and therefore the conversion of input values near the reference voltage might contain some uncertainty. In order to quantify this "transition voltage" range, it is not sufficient to measure the reference noise, as any extra effect occurring inside the comparator itself or lack of knowledge of the performance of the device would be neglected. We emphasise this measurement makes no assumptions at all about the circuit, making the measurement outcome ultimately transparent. As shown in Fig. S4, the measurement setup is as follows: we use a 3 dB splitter (Mini-Circuits ZFRSC-183-S+) after the photodetector to get two copies of the output analog random amplitude $v_{1,2}$. We send v_1 to the comparator input and v_2 to the oscilloscope, which also records the output of the first flip-flop, i.e., the comparator output latched exactly at the same point the oscilloscope is sampling. In order to match the sampling points of the oscilloscope and the flip-flop, we sweep the sampling point of the oscilloscope in steps of 10 ps until we find the best agreement, see Fig. S3 (left). Note that having the oscilloscope sampling at exactly the same point as the flip-flop is critical for this measurement. The analysis shows a small but finite overlap between range of input values giving low versus high output values near the threshold voltage. Computing the frequency of high and low outputs for each input value v_2, we obtain the conversion probabilities $P(d = x|v_2)$ for $x \in \{0, 1\}$.

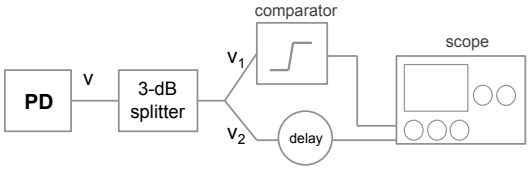

FIG. S4. Schematic for the characterization of the comparator chip. The signal from the photodetector (PD) is split in two by means of a 3-dB splitter. One of the signals is sent to the comparator and the other to the scope, and delayed using the skew function of the scope. The sampling point of the comparator can be tuned with 10 ps resolution. By capturing the digital signal output of the comparator and the analog signal recorded by the scope, we can measure the performance of the comparator, e.g. the probability of the output conditioned on the input.

The measurement suffers from some limitations. In an ideal scenario, we would need (i) the two outputs of the 3-dB splitter to be identical and (ii) the sampling point of the oscilloscope and flip-flop too. In practice, unfortunately, (i) the two outputs of the 3-dB splitter are not identical but their difference follows a Gaussian distribution with 0 mean and 1.32 mV rms deviation, and (ii) the timing precision is limited to 10 ps. Note that the presence of both limitations are conservative from the measurement point of view, i.e., the real error will always be smaller than the measured error. The narrowest observed transition is depicted in Fig. S3 (right), and shows an r.m.s. width of 8.4 mV. Considering that the oscilloscope noise is 3.4 mV rms, we can place an upper limit: $\sigma_{v_{\rm ref}} \leq 7.7\,\text{mV}$.

UPPER AND LOWER BOUNDS ON THE FRESHNESS TIME

As illustrated in Fig. 1 of the main article, the window for randomness generation is bracketed on the early side by the requirement for space-like separation from the distant detection, and on the late side by the requirement for space-like separation from the pair generation. Ensuring the random events fall in this window requires both upper and lower bounds on the freshness time. We measure timing of relevant events using a differential probe and a 20 GSa/s real time oscilloscope (Agilent Infinitum MSO9404A). As shown in Fig. S5, we measure three delays in the circuit: (i) t_1: from the modulation of the laser, measured directly on the pins of the laser, to the output of the photodetector, (ii) t_2: from the output of the photodetector to the input of the XOR gate, and (iii) t_3: from the input of the XOR gate to the CML output connector. We split in three intervals for traceability of the signal while travelling through the electronics.

All delays are quantified by capturing traces and using cursors to identify the zero-crossing times of relevant edges. As shown in Fig. S5, for each transition we identify a best guess $t_i^{(\text{best})}$ limited by the uncertainty of the interpolation between 50 ps samples. We find $t_1^{(\text{best})} = 7.82$ ns, $t_2^{(\text{best})} = 1.16$, and $t_3^{(\text{best})} = 1.55$ ns.

Measuring intervals on the oscilloscope can give a systematic error, which we now bound. In light of the $\sim 100 - 150$ ps 10%-90% rise and fall times of the transitions, it is extremely improbable that we misjudge the location of the edge by 100 ps or more, as it would mean placing the best guess outside of the transition region. As illustrated in Fig. S6, we calculate upper and lower bounds for the transition times as $t_i^{(\text{ub})} = t_i^{(\text{best})} + 100$ ps and $t_i^{(\text{lb})} = t_i^{(\text{best})} - 100$ ps, respectively. Combining the three measured intervals we find $\tau_{\rm f}^{(\text{ub,sys})} = \sum_{i=1}^{3} t_i^{(\text{ub})} = 10.87$ ns and $\tau_{\rm f}^{(\text{lb,sys})} = \sum_{i=1}^{3} t_i^{(\text{lb})} = 10.21$ ns. Fig. S7 shows the sequence of measurements performed.

The jitter of the signal plus the jitter of the oscilloscope is quantified by accumulating 10^7 traces using the persistence mode of the oscilloscope. Measuring the rising edge of the signal at the photodiode output (as in Fig. S2, e.g.), we observe that all of the traces fall within a 125 ps temporal window; there is no recorded value outside of this window. Making the hypothesis that events outside of this window occur with probability at least 1.4×10^{-6}, we expect ≥ 14 events on average outside of this window. A Poisson distribution with this mean predicts our observed zero events with probability 8×10^{-7}. We can thus reject the hypothesis and assign a $p < 1.4 \times 10^{-6}$ confidence to the 125 ps window for the measured zero-crossings. We note that this does not count the contribution of the oscilloscope to the jitter and is thus conservative. The freshness time combines intervals from three cascaded measurements. Adding three such windows, the conservatively estimated window for the full process is $3 \times 125 = 375$ ps. Half of this, 187.5 ps, can be assigned to the upper limit, and half to the lower. To have a round number, we define the jitter bound to be $\tau_{\rm jitt} = 200$ ps.

The lower and upper bounds for the freshness time of a single bit, including statistical and systematic errors, conservatively estimated, are then

$$\tau_{\rm f}^{(\text{lb,sys})} - \tau_{\rm jitt} < \tau_{\rm f} < \tau_{\rm f}^{(\text{ub,sys})} + \tau_{\rm jitt} \qquad \text{(S11)}$$

$$10.01\ \text{ns} < \tau_{\rm f} < 11.07\ \text{ns} \qquad \text{(S12)}$$

We also observe that the rms width of the jitter is $\sigma_{\rm jitt} < 40$ ps. If the jitter can be considered normally distributed, $\tau_{\rm jitt}$ is then at least $5\sigma_{\rm jitt}$, and could be made considerably stronger, e.g. at least $10\sigma_{\rm jitt}$, with only a 200 ps broadening of the bounds given above.

The operating frequency of the quantum random number generator is nominally 200 MHz, which corresponds to a cycle time of 5 ns. The clock of the RNG can be de-

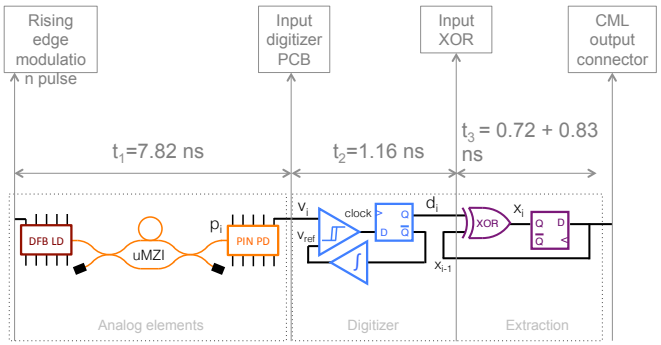

FIG. S5. (color online) Schematic of the random number generator with timing information as directly measured by a 20 GSa/s real time oscilloscope. t_1 describes the time from rising edge of the current pulse until rising edge of the PIN PD output pulse, t_2 describes the time from the PIN PD pulse to the input of the XOR gate, and t_3 the time from the input of the XOR gate to the output SMA port.

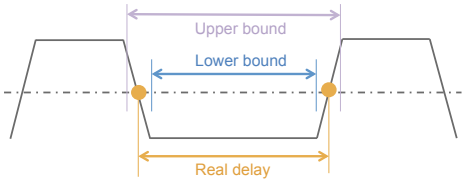

FIG. S6. (color online) For a given captured trace, the best-guess estimate for the zero-crossing time is found by interpolation between the 50 ps samples of the 20 Gsps oscilloscope, with systematic uncertainty (orange circles). To obtain the upper bound (purple) and lower bound (orange) for the zero-crossing time, we add or subtract the half-width of the transition. Statistical uncertainty is accounted separately from the statistics of 10^7 traces collected in the oscilloscope's persistence mode.

rived from an external reference via a phase-locked loop, or internally generated from a quartz oscillator (Analog Devices AD9522/PCBZ), in either case introducing a timing uncertainty that is nominally < 1 ps and negligible on the scale of the other uncertainties. The freshness time for k events is therefore given by

$$\tau_{\rm f}^{(k)} = \tau_{\rm f} + 5 \times (k-1) \text{ ns}. \tag{S13}$$

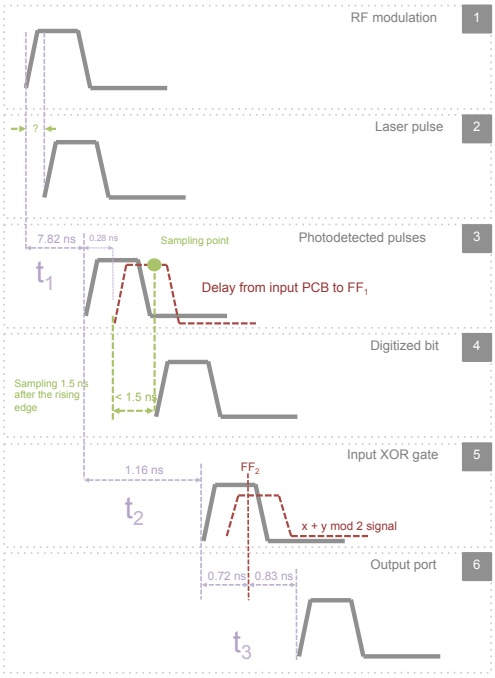

FIG. S7. (color online) Details of the sequence of measurements to obtain the freshness time. We start by measuring the modulation pulse directly at the laser diode pins (1). The LD generates an optical pulse some time after the modulation pulse is applied (2). Then, the signal is photodetected and measured (3). The difference between (1) and (3) gives t_1. Then, the electrical pulse is digitised by a fast comparator and latched by a first flip-flop (4), and then propagates until the input of an XOR gate (5). At the input of the XOR gate we can tap the signal and measure the arrival time. The difference between (3) and (5) gives t_2. Finally, the signal goes into the randomness extraction circuitry (see main text for more details), and propagates to the output, where we measure the arrival time again (6). The difference between (5) and (6) gives t_3.

STATISTICAL TESTING

A fast digitizer (Acquiris U1084A) is used to acquire the RNG output for statistical testing. Because of memory limitations, data is acquired in runs of ≈ 8 Mb each and concatenated.

Sample autocorrelation

The two-point autocorrelation is $\Gamma_x(k) \equiv \langle x_i x_{i+k}\rangle - \langle x\rangle^2$, where k is the correlation distance. To obtain experimental autocorrelations from a sample of N bits as in Fig. 5,we compute the unbiased estimator for $\Gamma_x(k)$

$$\hat{\Gamma}_x(k) \equiv \frac{1}{N-k}\sum_{i=0}^{N-k} x_i x_{i+k} - \left[\frac{1}{N}\sum_{i=0}^{N} x_i\right]^2. \tag{S14}$$

For a perfect coin and large N, the estimator has rms statistical uncertainty $\sigma_{\hat{\Gamma}} = 1/4\sqrt{N}$.

Considering the output of the RE, with $x_i = x_{i-1} \oplus d_i$, where d_i are raw bits, we note that x can be described as a symmetric two-state machine that changes state whenever $d = 1$. This symmetry guarantees the long-time average $\langle x\rangle = 1/2$, except if d is deterministic.

We estimate $\Gamma_x(k)$ when d has bounded predictability, i.e. $\max[P(d_i = 0), P(d_i = 1)] \leq \frac{1}{2}(1+\epsilon_{\max})$ for all i. We note that $x_i x_{i+k}$ is only nonzero for $x_i = x_{i+k} = 1$, so we can evaluate $\langle x_i x_{i+k}\rangle$ as the probability $P(x_i = 1) = \frac{1}{2}$ times the conditional probability $P(x_{i+k} = 1 | x_i = 1)$. This latter is the probability of an even number of $d = 1$ raw bits between i and $i+k$. Subject to the bound, this is maximized when $P(d_i = 0) = \frac{1}{2}(1+\epsilon_{\max})$ for all i. Counting the possible ways to have an even number, where n is the number of bits with $d = 1$, we find

$$\begin{aligned}\Gamma_x(k) &= \langle x_i x_{i+k}\rangle - \langle x\rangle^2 \\ &\leq \frac{1}{2^{k+1}}\sum_{n=0}^{\lfloor k/2\rfloor}\binom{k}{2n}(1-\epsilon_{\max})^{2n}(1+\epsilon_{\max})^{k-2n} - \frac{1}{4} \\ &= \frac{1}{4}\epsilon_{\max}^k. \end{aligned} \tag{S15}$$

Statistical test batteries

Several statistical batteries are used to test the quality of the output, including the TestU01 Alphabit battery

[35], the NIST SP800-22 battery [34], and the Dieharder battery [33]. The results are consistent with ideal randomness for $k = 3$ and above.

Due to the high output rate of the RNG, testing was limited by computation speed for the various tests. In this regard Alphabit has a significant advantage, as it was designed for testing physical RNGs, without the more computationally-intensive tests used for pseudo random number generators. For example, testing a 1.5 Gb sequence with the NIST battery takes more than 3 hours on a desktop computer whereas the Alphabit battery takes one minute.

Dieharder tests

We ran the dieharder test with default settings. Results are shown in Table S-II.

NIST SP800-22

As per NIST recommendations [34], we use $m = 1500$ sequences of 1 Mb each to assess the random numbers generated by the device. The tested sequences pass both the proportion and uniformity of the p-values assessments. See Fig. S8.

Test U01 Alphabit battery

We use the Alphabit battery following two testing strategies: (i) test many different sequences of a relatively small size, e.g. 300 files of 1 Gb each, and (ii) testing very long sequences, e.g. one file containing 1 Tb. Using (i), we can quantify how often the generator fails the Alphabit battery. This is important because an ideal random number generator should fail with around 2% probability. With (ii) we can test for weaker correlations/anomalies, below the statistical uncertainty of strategy (i).

Results for strategy (i) are depicted in Fig. (S10). We tested 1 Mb for k=1, 120 Mb for k=2, 500 Mb for k=3, and 1 Gb for k=4. In each case we test 300 sequences. For strategy (ii) we tested a single 1 Tb file, two 500 Gb files, one 80 Gb file, and two 64 Gbfile for $k = 3$ and all tests were passed. We followed the same criterion for evaluating the results as in [42] in which regularities in commercial RNG systems were found for 64 Gb and above.

[36] G. Agrawal, IEEE J. Quantum Electron. **26**, 1901 (1990).
[37] M. Scully and M. Zubairy, *Quantum Optics* (Cambridge University Press, Cambridge, England, 1997).
[38] H. E. Rauch, C. T. Striebel, and F. Tung, AIAA J. **3**, 1445 (1965).
[39] A. Gelb, *Applied Optimal Estimation* (MIT Press, Cambridge, MA, 1974).
[40] A. Einstein, Deutsche Physikalische Gesellschaft Verhandlungen **18**, 318 (1916).
[41] J.-Å. Larsson, J. Phys. A **47**, 424003 (2014).
[42] K. S. Jakobsson, Master's thesis, Linköpings Universitet (2014)

TABLE S-II. Summary of results for the entire Dieharder battery for k=3. In column `n tuple`, the notation $1 \ldots x$ indicates that the test was repeated with the `n tuple` setting covering this range. As shown, the diehard parking lot test showed a weak value, i.e., an inconclusive result, in the initial run. For an ideal source, it is expected that ~ 1 weak value will appear in any given full test run of the suite. As recommended, we re-run the weak test with the option `-Y`, which increases `p samples` until a clear result (passed or failed) emerges. The test was passed.

test name	**n tuple**	**t samples**	**p samples**	p**-value**	**Assessment**
diehard birthdays	0	100	100	0.86914871	PASSED
diehard operm5	0	1000000	100	0.62710352	PASSED
diehard rank 32x32	0	40000	100	0.49138373	PASSED
diehard rank 6x8	0	100000	100	0.46907910	PASSED
diehard bitstream	0	2097152	100	0.94865200	PASSED
diehard opso	0	2097152	100	0.41217400	PASSED
diehard oqso	0	2097152	100	0.49075022	PASSED
diehard dna	0	2097152	100	0.78245172	PASSED
diehard count 1sstr	0	256000	100	0.59545874	PASSED
diehard count 1sbyt	0	256000	100	0.44152512	PASSED
diehard parking lot	0	12000	100	0.99601277	WEAK
diehard 2dsphere	2	8000	100	0.93564067	PASSED
diehard 3dsphere	3	4000	100	0.94103902	PASSED
diehard squeeze	0	100000	100	0.58677274	PASSED
diehard sums	0	100	100	0.76102130	PASSED
diehard runs	0	100000	100	0.11968119	PASSED
diehard runs	0	100000	100	0.51728489	PASSED
diehard craps	0	200000	100	0.86247445	PASSED
diehard craps	0	200000	100	0.97041678	PASSED
marsaglia tsang gcd	0	10000000	100	0.68088920	PASSED
marsaglia tsang gcd	0	10000000	100	0.20851577	PASSED
sts monobit	1	100000	100	0.15319982	PASSED
sts runs	2	100000	100	0.80047009	PASSED
sts serial	$1 \ldots 16$	100000	100	0.11969309 – 0.98698338	PASSED
rgb bitdist	$1 \ldots 12$	100000	100	0.07588499 – 0.96001991	PASSED
rgb minimum distance	2	10000	1000	0.19433266	PASSED
rgb minimum distance	3	10000	1000	0.45871020	PASSED
rgb minimum distance	4	10000	1000	0.53968425	PASSED
rgb minimum distance	5	10000	1000	0.82922021	PASSED
rgb permutations	2	100000	100	0.59213936	PASSED
rgb permutations	3	100000	100	0.77283570	PASSED
rgb permutations	4	100000	100	0.98499840	PASSED
rgb permutations	5	100000	100	0.47842741	PASSED
rgb lagged sum	$0 \ldots 32$	1000000	100	0.11439306 – 0.99357917	PASSED
rgb kstest test	0	10000	1000	0.81377150	PASSED
dab bytedistrib	0	51200000	1	0.45890415	PASSED
dab dct	256	50000	1	0.80469155	PASSED
dab filltree	32	15000000	1	0.22779072	PASSED
dab filltree	32	15000000	1	0.61178536	PASSED
dab filltree2	0	5000000	1	0.18837842	PASSED
dab filltree2	1	5000000	1	0.49022984	PASSED
dab monobit2	12	65000000	1	0.43688754	PASSED

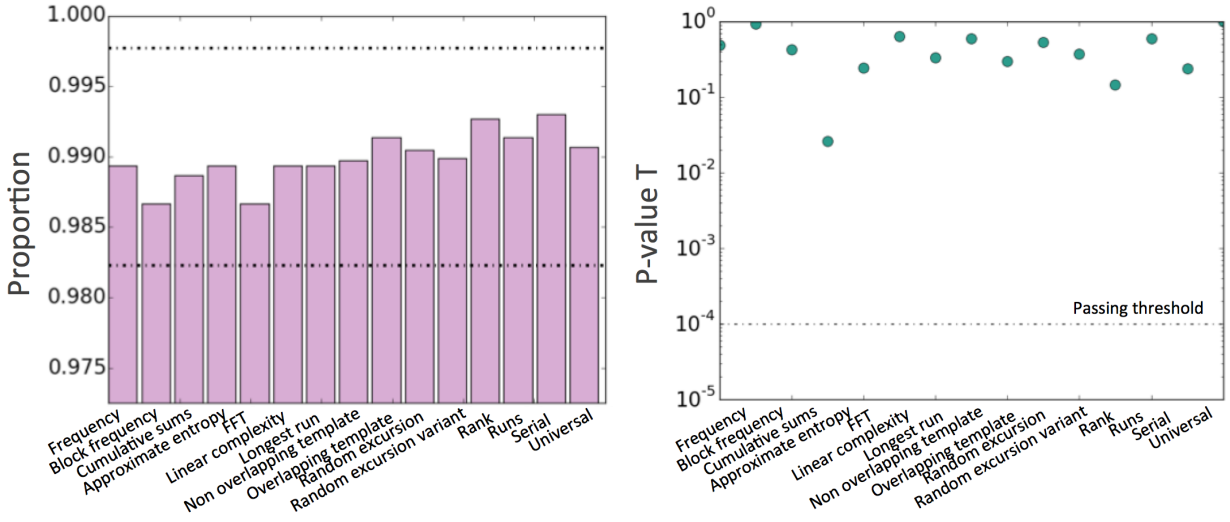

FIG. S8. Summary of results for the NIST SP800-22 battery. The top figure depicts the results for the proportion test, i.e. evaluating how many times each test passes or fails (note that even a perfect random number generated is supposed to fail some tests from time to time). Dot-dash lines indicate limits of acceptable behavior as recommended by NIST SP800-22. The bottom picture represents the uniformity of the p-values assessment, which mainly quantifies how the outcomes of each statistic (p-values) are distributed along the (0,1) range. Both quantities satisfy the statistics for a perfect random number generator.

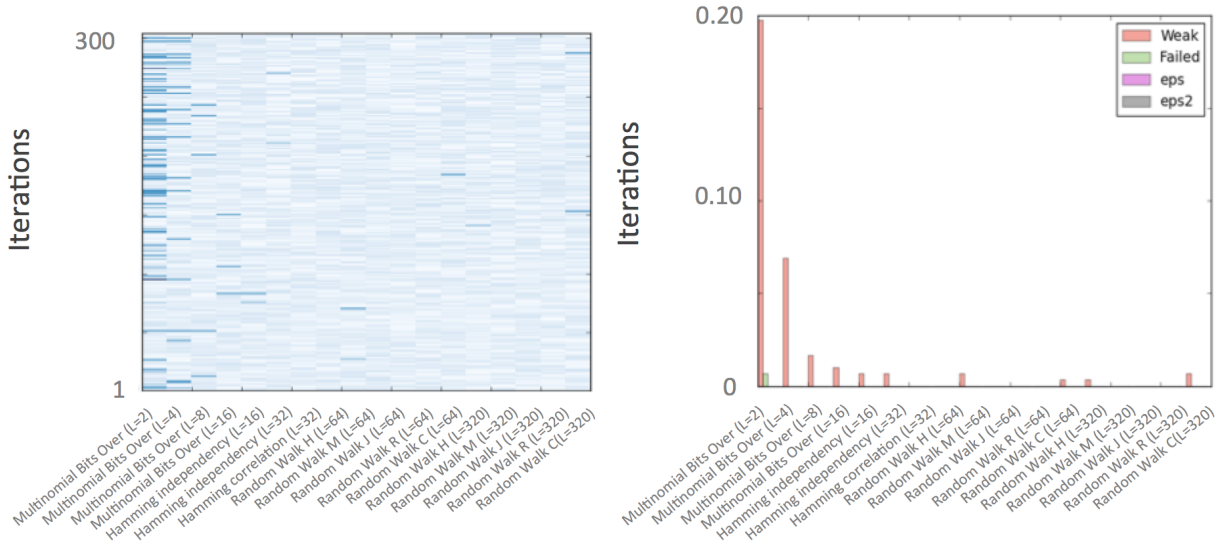

FIG. S9. Statistical tests for $k = 1$. (Left) represents the results for each of the 17 tests (horizontal axis) in the Alphabit battery for 300 iterations (vertical axis). A dark blue square represents a weak value is observed for that particular iteration and test. For each test, a p-value p is computed and its deviation from zero or one $\Delta_{(p)} = \min\{1-p, p\}$ is measured. If $\Delta_{(p)} \geq 10^{-2}$ the test is considered to pass that statistic. In contrast, a test is considered inconclusive or weak when its deviation is $10^{-2} > \Delta_{(p)} \geq 10^{-6}$, failed when $10^{-6} > \Delta_{(p)} \geq 10^{-15}$, "eps," which implies catastrophic failure, when $10^{-15} > \Delta_{(p)} \geq 10^{-300}$, and "eps2" when $\Delta_{(p)} < 10^{-300}$. (Right) shows the frequency of obtaining weak, failed, eps, and eps2 p-values for each of the 17 tests of the alphabit battery. As shown, for k=1 the alphabit battery is able to find statistically significant anomalies, indicating the need for randomness extraction.

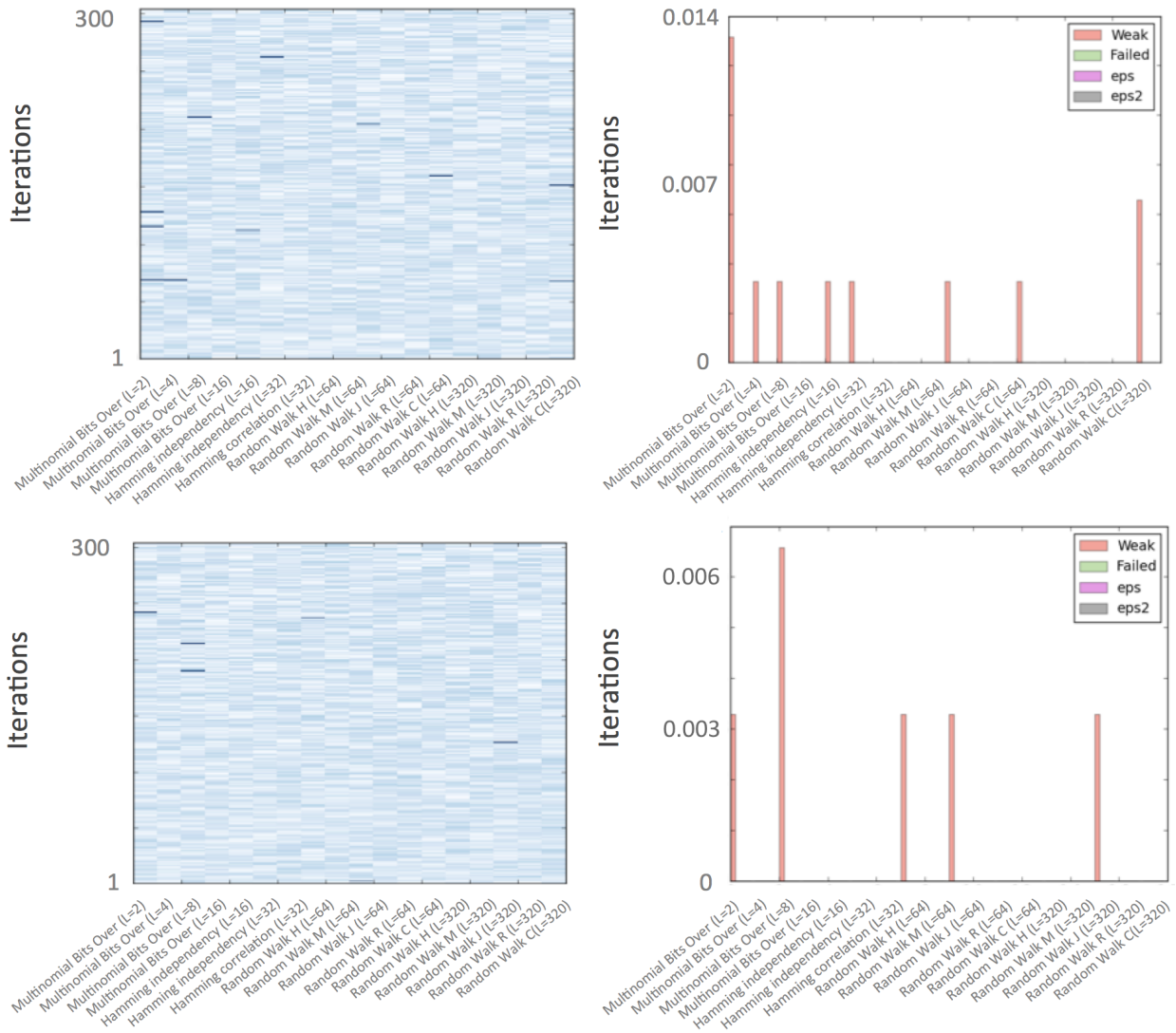

FIG. S10. (Rows 1 and 2 correspond to the statistical tests for $k = 2$ and $k = 3$, respectively. (Left column pictures) represent the results for each of the 17 tests (horizontal axis) in the Alphabit battery for 300 iterations (vertical axis). A dark blue square represents a weak value is observed for that particular iteration and test. For each test, a p-value p is computed and its deviation from zero or one $\Delta_{(p)} = \min\{1 - p, p\}$ is measured. If $\Delta_{(p)} \geq 10^{-2}$ the test is considered to pass that statistic. In contrast, a test is considered inconclusive or weak when its deviation is $10^{-2} > \Delta_{(p)} \geq 10^{-6}$, failed when $10^{-6} > \Delta_{(p)} \geq 10^{-15}$, "eps," which implies catastrophic failure, when $10^{-15} > \Delta_{(p)} \geq 10^{-300}$, and "eps2" when $\Delta_{(p)} < 10^{-300}$. (Right column pictures) show the frequency of obtaining weak, failed, eps, and eps2 p-values for each of the 17 tests of the alphabit battery. For k=2, the results are statistically acceptable, but appear borderline. For k=3,4 (the k=4 case is not shown), the failure rate is consistent with an ideal randomness source.